

## FedAvg

- 相较于传统SGD每个client计算完梯度就发给server，FedAvg计算完会更新局部参数，迭代多次，再发给server。
- **FedAvg是典型的以计算换通信策略**，在相同通信次数下FedAvg更快收敛，client相同计算量（epoch来衡量）请款下，传统SGD收敛更快。而联邦学习中计算代价小，通信代价大。

## FedProx

- 相较于FedAvg提供了一个收敛更快效果更好的算法
- FedProx作者还探究了统计异质性和系统异质性之间的相互作用，并认为系统异质性产生的**掉队者 (stragglers)** 以及掉队者发往server的带偏差的参数信息会进一步增加统计异质性，最终影响收敛。因此，作者提出在client的优化目标函数中增加一个**近端项**，这样可以使优化算法更加稳定，最终使得FedProx在统计异质性下也收敛更快。
- 近端项的存在考虑最小化训练后获得的权重与服务器获得的权重的差异，使得 $w$ 不与原始值偏离太多，限制了因为数据异质性导致的模型偏移
- server操作和FedAvg相同
- client不再执行E轮SGD，改为求解带近端项的优化问题

## FedAvg+

- 个性化联邦学习，不求训练出全局模型，使每个节点训练各不相同的模型，采用**模型不可知的元学习**

元学习学习在给定小样本实例的条件下进行自适应，可以优化在异构任务上的表现，元学习步骤：

1. meta training 训练初始化/元模型
  2. meta testing 使初始化模型在特定的任务下完成自适应
- 将全局的已经训练的模型视为初始化模型，局部模型视为个性化模型
  - 算法步骤：
    1. FedAvg得到初始化模型
    2. 用FedAvg变种算法对模型微调
    3. 对client进行个性化操作，采用和训练期间相同优化器

## Clustered FL

- 针对数据Non-IID导致的局部最优
- 每个节点训练不同模型，让节点训练的过程进行**知识共享（并不共享参数，而共享参数的变化量）**，无需另设初始化模型
- 训练过程中把参数相似的节点划分为同一个任务簇，共享参数的变化量 $g$ ，完成知识共享和相似节点相互促进目的

聚类联邦学习算法每轮通信描述：

1. 第k个client从server接收变化量更改本地参数，执行局部epoch的SGD，将训练出来的变化量发往server
2. server接收到K个client发来的变化量，对每个簇求解**簇内平均参数变化**，根据不同节点的参数变化了的余弦距离重新划分聚类簇

## pFedMe

---

- 用Moreau-Yosida正则化作为client的正则损失函数，达到更快的收敛速度
- 将个性化模型与全局模型同时进行优化求解
- 该方法按照与标准FedAvg相似的方法来更新全局模型（多了个一阶指数平滑），不过会以更低的复杂度来对个性化模型进行优化。

## FedEM

---

- 从优化算法上提高联邦学习精度
- 基于client节点数据满足混合分布假设，使得每个client节点训练由M个子模型集合所得模型
- 针对混合分布假设采用EM算法做参数估计，提高模型整体精度。